

Development of the Federal Statistical System Public Opinion Survey

Jennifer Hunter Childs¹, Stephanie Willson², Shelly Wilkie Martinez,³
 Laura Rasmussen⁴, Monica Wroblewski¹

¹ U.S. Census Bureau, 4600 Silver Hill Rd. Suitland, Md. 20746

² National Center for Health Statistics, 3311 Toledo Rd. Hyattsville, MD 20782

³ Office of Management and Budget, 725 17th St, NW Washington, DC 20503

⁴ Internal Revenue Service, 1111 Constitution Ave. Northwest DC 20004

Abstract

The U.S. Census Bureau is collaborating with other federal statistical agencies to understand public trust in official statistics in the United States. Over the next two years, this interagency group is commissioning a public opinion survey of attitudes toward statistics produced by the federal government. The study looks at trust in the federal statistical system, the credibility of federal statistics, and attitudes toward and knowledge of statistical uses of administrative records. The research follows similar efforts in European countries that have traced public views of official statistics. The design of the study will also allow us to make comparisons between attitudes observed in the U.S. and those measured in Europe and observe how current events correlate with public perception towards the federal statistical system. This paper describes the development and cognitive testing of the questionnaire employed to measure trust in official statistics in the U.S. Results preliminary information about the structure of public opinion towards these topics in the U.S. and informed the selection of questions for a field pretest.

Key Words: Federal Statistics, Trust, Credibility, Administrative Records, Cognitive Interviewing, Questionnaire Development

1. Background¹

From December 2009 through April 2010, the U.S. Census Bureau hired a private survey firm to conduct a nightly survey of public attitudes toward the 2010 Census, public awareness of 2010 Census advertising, and intent to mail back their Census forms. The nationally representative sample of 200 respondents per night was rolled up into seven-day moving estimates that provided nearly immediate feedback on public reaction to national events that might influence perceptions of the 2010 Census, and on the reach and reception of 2010 Census communications campaign messaging. The Census Bureau used this feedback to make communication campaign decisions during the 2010 Census that contributed to achieving a mail-back participation rate of 74 percent, despite increased vacancy rates due to economic downturn, increased public skepticism about the role of the Federal Government, and a general decline in survey response rates during the decade that affected both the public and private sectors. Details of this study are chronicled in Miller and Waleko (2011).

Moving forward, the Federal Statistical System must find ways to reverse the decline in response rates for their ongoing surveys or face both increasing operational costs and declines in data quality. It also must find more comprehensive ways to reuse data already collected for programmatic and other purposes if it is to meet ever-expanding needs for information in a resource-constrained environment. Specifically, the member agencies of the Interagency Council on Statistical Policy (ICSP) have expressed an interest in collecting data to assess attitudes, beliefs, and concerns the public may have regarding its trust (or confidence) in federal statistics and in the collection of statistical information by the federal government from the public, as well as attitudes toward and knowledge of the statistical uses of administrative records. The data will also help us to understand how current events appear to impact public perception

¹ Disclaimer: This paper is released to inform interested parties of research and to encourage discussion of work in progress. Any views expressed on methodological or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau or the other agencies mentioned within.

towards federal statistics. Ultimately, this improved understanding of public perceptions will guide FSS communications with the public and future plans for data collection.

A subgroup of ICSP members spearheaded this effort. They include the heads of the Census Bureau, the National Agricultural Statistics Service, the National Center for Health Statistics, the Economic Research Service, the Statistics of Income Division (IRS), and the Statistical and Science Policy Office of the Office of Management and Budget. We refer to this group of agency heads as the Federal Statistical System (FSS) Team. These agencies put forth working group members who will be referred to as the FSS Working Group. It was this working group that developed and tested questions for this effort.

In a similar endeavor, an Organization for Economic Co-operation and Development (OECD) working group developed a survey for measuring trust in official statistics that was cognitively tested in six of the member countries (Brackfield 2011). The goal of that undertaking was to produce a model survey questionnaire that could be made available internationally to be used comparably in different countries. Many national statistical offices are recognizing the critical role of public trust and robust communication to ensure high quality data, particularly in an era of constrained resources. This international effort recognized that rather than relying on anecdote or no evidence at all, having objective, quantifiable information about public attitudes is needed to inform decision-making. Unfortunately, a 2010 National Center for Health Statistics (NCHS) cognitive study revealed that the OECD questions, which depend to a large extent on the concepts of a single centralized National Statistical Office that produces most all of a country's "official" statistics, are inadequately understood by U.S. respondents (Willson 2010) and therefore, would be unable to sufficiently measure trust in the FSS in the United States.

In 2011, the Census Bureau's Communications Directorate conducted the second iteration of the Census Barriers, Attitudes, and Motivators Survey (CBAMS II) as a follow-up to the original CBAMS conducted prior to the 2010 Census in 2008. CBAMS was conducted to gain an in-depth understanding of the public's opinions about the 2010 Census, with the specific intention to understand those who have negative attitudes toward the Census and the government more generally or those who are unaware/lack extensive knowledge of the Census. CBAMS II provided a post-2010 Census measurement of the same issues; this time, though, the questionnaire also included questions on the use of administrative records for the decennial census, which relates to this study. In CBAMS II, respondents were experimentally divided into three groups in order to test their views of administrative records use as a means of (1) reducing Census (government) costs, (2) reducing respondent burden or (3) as simply an alternative option to a self-response (the control group). From this research, the study found that both arguments of reducing cost (when citing a \$10 billion Census price tag) and of alleviating respondent burden increased public support of administrative records usage, though the cost reduction frame was more powerful (Wroblewski, Bates and Pascale, 2012; Conrey, ZuWallack, and Locke, 2011). Additionally, the CBAMS II found that some administrative records are less sensitive than others are. People were more comfortable with obtaining one's name, date of birth, gender and race from tax returns (50%), or other government records such as unemployment or social security (45%); whereas they were much less in favor of the Census obtaining credit bureau data (25%) or medical records (22%) for use in a decennial census. Further, in the study, most people (65%) expressed unwillingness to allow the Census Bureau to use SSNs to obtain sex, age, date of birth and race information from other government agencies.

Other research has suggested the importance of providing a context for answering such questions, and CBAMS II, like many telephone surveys, afforded limited opportunity to provide such context. Previous research presents a somewhat conflicting picture of the topic – on one hand, public favorability toward the use of administrative records looks to be declining (Singer, Bates, Van Hoewyk, 2011). On the other hand, a recent study of public willingness to grant informed consent to record use paints a more optimistic picture (Pascale, 2011).

Building upon this research, the FSS Working Group hopes to understand the interrelationships between knowledge, trust and opinions towards the use of administrative records to help guide future policy and

communication efforts. As such, the FSS Working Group began the development of new questions that will measure these concepts of trust in a language familiar and understandable to respondents in the U.S. This development is described in the next section.

2. Questionnaire Development

2.1. Trust in Statistics

As mentioned, the OECD developed a survey measuring trust in official statistics, but cognitive testing of these questions by NCHS in 2010 revealed that the survey questions did an inadequate job of conveying intent to the respondents. In fact, most respondents did not understand the questions as asking about their knowledge and opinion of official statistics at all. Instead, a common interpretation was that the questions were asking about respondent knowledge and opinion of the *substantive topic itself*, for example unemployment or crime (Willson, 2010). This problem is evident in the fact that respondents were more likely to interpret the unemployment and crime rate questions in terms of their experience with being unemployed or with being victims of crime. Interestingly, this pattern did not vary greatly by educational attainment: those with higher and lower levels of education were both likely to misunderstand the questions. Given this finding, the FSS Working Group sought to build upon the theoretical constructs and previous research on this subject (Fellegi 2004; OECD Working Group 2011; Wilson et.al. 2011) in designing and administering a version of this survey that might adequately measure U.S. public opinion of the FSS. The OECD questions that worked well in the NCHS cognitive test were included in the draft FSS survey for further cognitive testing. We modeled additional questions after other research intended to measure similar issues of trust and awareness in official statistics.

The FSS team focused on definitions of trust in statistical products and trust in statistical institutions derived from work by Ivan Fellegi (1996, 2004). Figure 1 shows his theoretical model and Figure 2 shows the components of trust as defined by Fellegi. The team focused on factors that would be familiar to the general population, who are not likely to be sophisticated data users with a high level of knowledge of the statistical system.

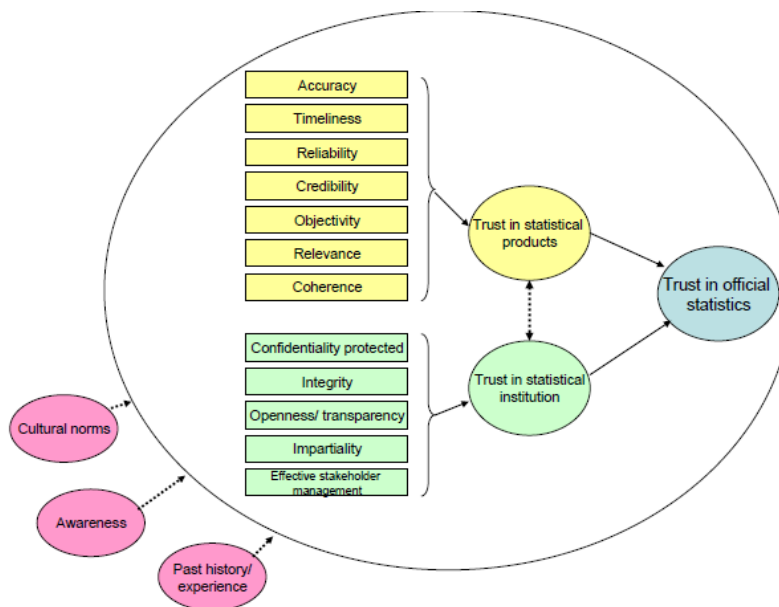


Figure 1. Fellegi's model of Trust in Official Statistics²

² Reproduced from OECD (2010).

Trust in Statistical Products

Accuracy- Accuracy refers to how well the statistical value presented in the data set matches up to the phenomena which it is trying to measure (VALIDITY).

Timeliness- In order for the public to trust that the statistical agency is producing un-manipulated, accurate information, the data must be released in a cost-efficient and punctual manner.

Reliability- Although the breadth of the data analysis conducted by government agencies may go beyond the capacity of other agencies, the data produced must include clear methodology and indication that the results produced are replicable—so that other agencies researching similar topics would come up with similar estimates to the data that are presented.

Credibility- Users rely on the reputation of the provider of the information to validate the data it produces. How credible the statistical products are determines the intrinsic value and usability of the information.

Objectivity- The agency must be seen as having no interest in producing particular figures or being linked to a partisan entity.

Relevance- It is important that the statistical product be seen to be meeting the needs of the nation, and not just those of the government. Relevance could be examined at national, local or personal levels.

Coherence- Across various measures and testing, the statistical products should be perceived as being derived from consistent methods and standards of rigor, regardless of where they were produced so that the data produced at any time can be trusted as being produced with the highest levels of statistical integrity.

Trust in Statistical Institution

Confidentiality Protected- The statistical agency requires a clear and visible mandate providing it with the authority to collect data for statistical purposes (and only statistical purposes) and the obligation to protect the confidentiality of individual responses.

Integrity- The reputation of the statistical agency is based upon a number of factors which cumulatively rely upon an ethic of consistent honest research. The nation should come to expect consistent, accurate, honest, non-partisan data from an organization with a clean record, or at least who has addressed previous events that may have tarnished their reputation.

Openness/Transparency- The level to which the statistical agency makes known the behind-the-scenes details of the data collecting process. This involves making the following information publically available: financial and political liaisons, statistical methodology and the data produced.

Impartiality- the extent to which the statistics are perceived to be objective and independent, unbiased and non-partisan (i.e. not subject to political interference).

Effective Stakeholder Management- Since the data produced by the statistical agency are supposed to serve the interests of the nation, the statistical agency must respond to the citizens of the nation to which they are responsible.

All definitions adapted from Fellegi (1996; 2004).

Figure 2. Fellegi's Constructs to Measure Trust

In addition to considering questions developed and tested by the OECD working group, the FSS working group considered questions used by the Office of National Statistics and the National Centre for Social Research in the United Kingdom and by the Eurobarometer. The group started from the premise that we needed to measure awareness of statistics and statistical institutions first, and then assess level of knowledge or data use, before proceeding to questions addressing trust. We consulted previous research that examined the U.S. public's knowledge of statistics (Curtin, 2007) and sought to create a questionnaire that would be comprehensible by the general population. The trust questionnaire created for cognitive testing, as well as the source of each question and the construct it attempts to measure, are found in Appendix A.

2.2. Administrative Records

The second overarching goal of this effort is to gauge public opinion towards the use of administrative records for statistical purposes. For this part of the questionnaire development, we reviewed previous

questions on the topic, mostly from the perspective of the Census Bureau (Miller and Walejko, 2010; Singer, Bates and Van Hoewyk, 2011; and Conrey, ZuWallack, and Locke, 2011).

We also considered work conducted internationally in this area. A study conducted by the ONS (2009) revealed that the public in the U.K. has varied degrees of knowledge about government agencies and their current levels of data sharing (to permit one agency to rely on another's administrative data). Over 50 percent of respondents were aware that no single central government database currently exists, but that separate databases are maintained by individual departments, though this knowledge varied by education, age and region. Overall the response received was supportive (approximately two-thirds in favor) of data sharing and the creation of a single central population database of U.K. residents. By including similar questions about knowledge of and attitudes about data sharing, the FSS may be able to take measures to increase awareness and/or produce quantitative evidence in support of altering current data sharing practices, which may enable the government to save costs and improve data quality.

2.3. Implications

We are interested in examining the interrelationships between knowledge, components of trust and credibility and opinions towards administrative records. Through the nightly survey, we hope to see how the public responds to current events and how different types of events differentially influence trust, credibility, and views towards administrative records use, and how this may vary by knowledge and demographic characteristics. These factors could be used for public opinion campaigns to increase public awareness of the FSS and confidence in administrative records use or to contradict any negative publicity that occurs in this domain.

2.4. Study Design

This study aims to better understand Americans' knowledge of and attitudes toward the FSS as a means to improve the agency's survey response rates and data quality. Based on the pretesting described in this paper and in Earp et al. (forthcoming), the Census Bureau has added 25 questions to a sample of cases in the ongoing Gallup Daily Tracking survey. This daily survey asks U.S. adults about various political, economic, and well-being topics.

The analytic goals of the FSS Public Opinion Survey (POS) are to:

- Examine the relationship between awareness of and trust in the federal statistical system and federal statistics in the United States;
- Explore knowledge about and attitudes towards the statistical uses of administrative records;
- Examine the relationship between trust in the statistical system and attitudes towards the statistical uses of administrative records;
- Observe how current events influence public perception towards the federal statistical system;
- Develop a time series of trust in the statistical system;
- Make comparisons between attitudes observed in the U.S. and those measured in Europe; and
- Inform a variety of internal management decisions, such as how to provide better information with the public to address any misperceptions or concerns raised by the survey results (see Miller and Walejko, 2010 for an example of how awareness activities could be targeted).

The survey methods for the Gallup Daily Tracking rely on computer-assisted telephone interviews, dual-frame sampling (which includes a listed landline frame combined with an RDD Cellphone frame³), and a

³ Survey Sampling, Inc. provides listed landline sample and RDD cell phone sample (consisting of all exchanges set aside for cell phones) in non-overlapping frames. The sample is stratified proportionately by census region.

random selection method for choosing respondents within the household.⁴ Additionally, the survey includes Spanish-language interviews for respondents who speak only Spanish, sample coverage in Alaska and Hawaii, and relies on a three-call design to reach respondents not contacted on the initial attempt. Nightly quotas exist to ensure that the un-weighted samples are proportionate by region. The data are weighted daily to compensate for disproportionalities in selection probabilities and non-response. The data are weighted to match targets from the Census Bureau by age, sex, region, gender, education, ethnicity, and race. With the inclusion of the cell phone-only households and the Spanish-language interviews, more than 90 percent of the U.S. adult population is represented in the sample. By comparison, typical landline-only methodologies represent less than 70 percent of the adult population. The Gallup Daily Tracker response rate has averaged 11 percent, using an established formula for calculating response (CASRO).

2.5. Rotating Questions

As mentioned previously, the FSS survey consists of 25 questions added to an ongoing daily data collection by the Gallup organization for a period of approximately 20 months. Currently 19 of the 25 questions are core questions and six are designated for rotation. Core questions will primarily focus on awareness of and attitudes towards federal statistics and federal statistical agencies with one or two focusing on attitudes towards the statistical use of administrative records. Core questions will be used to explore relationships among the concepts, develop a time series and measure any “shocks” to the system. Shocks could include any current events that influence awareness or attitudes towards topics being measured. They may include things like data breeches (public or private sector), elections, or any unanticipated news event that may alter public perception of federal statistics or federal statistical agencies. By having a continual data collection, we will be able to look for changes in public perception after any of these types of events occur or look for underlying causes when we see a change in the time series.

Up to 20 times during the data collection, roughly monthly, a maximum of six questions may be rotated in the survey. Rotating questions will be used for three distinct purposes:

- First, experiments will be planned for questions surrounding public opinion of administrative records use for statistical purposes. Topics of experiments will include the public perception of the quality of such records, the public perception of privacy and confidentiality implications of their use, and differentiation between the types of administrative records and methods of use. These experiments will be conducted using different frames around similar questions, varying the types of records mentioned and the methods of use in the question, willingness-to-pay/stated preference questions, and so on.
- Second, rotating questions will be used around *known, planned events* to gauge awareness of those events and opinions about the relationship (if any) between those events and the federal statistical system. Examples of planned events are the presidential election, release of particular statistics, and any pre-planned public relations campaigns.
- Third, we may wish to add rotating questions very quickly *after an unanticipated* event to gauge awareness of those events and opinions about the relationship (if any) between those events and the federal statistical system. These could be events like a data breach (public or private sector), political scandal, or any other unanticipated news event that may alter public perceptions about the federal statistical system.

2.6. Limitations

⁴ When calling a landline telephone, random selection is used to choose respondents within a household based on the most recent birthday. Cell phones are treated as *personal devices*, where the individual answering is considered to be the respondent.

Certain aspects of our research design introduce limitations to this study and will necessitate some caution in interpreting and understanding the results.

Although the Gallup Daily Tracking is portrayed as being nationally representative, it does not meet Census Bureau quality standards for dissemination and is not intended for use as precise national estimates or distribution as a Census Bureau data product. The Census Bureau and other member agencies of the ICSP will use the results from this survey to monitor awareness and attitudes, as an indicator of the impact of potential negative events, and as an indicator of potential changes in communication campaigns. The CBAMS II study that surrounded the 2010 Census illustrates how these data can be useful for these types of decisions (Conrey, ZuWallack, and Locke, 2011). Conrey, ZuWallack, and Locke (2011) also demonstrate the usefulness of examining differences in public opinion among different demographics, such as race and age. Like the 2010 Census study, data from this research will be included in research reports with the understanding that the data were produced for strategic and tactical decision-making and not for official estimates.

The questions included in this study are administered as part of a larger nightly survey. As a result, questions asked before the FSS-specific ones may have an impact upon respondents' answers. The influence of the polling company's identity or interviewer administration may also influence the respondents' answers.

Those who choose to participate in a telephone poll (or any poll for that matter) may not be fully representative of the U.S. population. This may especially be a problem given that one of the central goals of this research is to determine trust in the FSS. Those respondents who refuse to participate in the phone survey, and therefore would not be included in our sample data and analysis, may be a population who particularly distrusts and/or is skeptical of such statistical agencies and/or the government. To address this in part, Gallup is conducting a nonresponse bias study designed by the Working Group. Furthermore, we are not able to assess the public opinion of those people who either do not have a phone or have an unlisted phone number. As such, we may miss out on the opinions of those without a phone (the homeless, etc.) or those who have an unlisted number, particularly younger populations especially who only have a cellphone and request to be unlisted or placed on a "do not call list" or those living in an institutions such as a nursing home.

Finally, this study will only address those interviews conducted in English and Spanish. The absence of the survey in other languages may exclude certain populations and their opinions regarding the FSS.

3. Cognitive Test

To develop the FSS survey, in 2011, the Census Bureau, NCHS, and IRS conducted 42 cognitive interviews of a purposive sample that was diverse in terms of age, race, gender, education and trust in the government (see Willson, 2012, for more information and question-level analysis). Interviews took place in a variety of locations, many outside of government buildings, in an effort to create an environment conducive to respondents expressing honest opinions, including distrust of government and government institutions. Interviews were conducted in the Washington D.C. and Atlanta Metro areas and were designed to last 60 minutes. A \$40 honorarium was given to respondents at the conclusion of the interview.

3.1. Methodology

Cognitive interviewing is a qualitative methodology that offers the ability to understand the interpretive process behind answers to survey questions. The style of cognitive interviewing utilized here involves respondent narrative and intensive follow-up verbal probing. Interviewers begin by administering the survey question, obtain an answer, and then "probe" the respondent for information with respect to the responses given. Interviewers can ask follow-up questions when contradictory information is given by the respondent as this may indicate a problem with the interpretation of the question itself. Probes are also

used to explore areas of concern within the instrument that have been pre-identified by the researchers. At the same time, interviewers have the option to explore unanticipated issues that may come up during the interview. The goal of this method is to understand what the respondent was thinking when answering and how they interpreted the meaning of the question. This information is used to help the researcher identify *which* questions and/or response categories are problematic. It also shows *why* and *how* questions are problematic, leading to informed strategies for improving question design in terms of maximizing construct validity.

3.2. Method of Analysis

Data analysis was conducted by the NCHS and used the grounded theory approach, which does *not* aim to test existing hypotheses, but instead generates explanations of response error and various interpretive patterns that are closely tied to the empirical data. The process of analysis is a constant comparison of data in three distinct steps. The first step occurs within the interview as the interviewer attempts to understand how one respondent has come to understand, process, and then answer a survey question. Basic response errors can be identified by comparing respondents' answers to the survey questions to the narrative they provide during the interview. When logical contradictions are evident between the narrative and the survey answer, the interviewer explores why these contradictions occurred. The second step in analysis occurs once the interview is over, and is a systematic comparison *across* all interviews. This level of comparative analysis reveals *patterns* in the way people answer survey questions. At this level, it is possible to identify the construct that is being captured by the survey question and illustrate the substantive meaning behind the survey statistic.

The final phase of analysis is a comparison of patterns across sub-groups, identifying whether particular *groups* of respondents interpret or process a question differently from other groups. At this level of analysis, that is, identifying patterned differences among subgroups, we begin to understand where potential for bias would occur in survey estimates.

4. Overview of Findings

The Challenge: Respondents' Knowledge of the Topic (or Lack Thereof)

The driving factor shaping the question-response process in these questions was respondents' lack of understanding and knowledge of the FSS in particular and statistical information in general. This is consistent with findings from the OECD international effort aimed at measuring attitudes about the general population's trust in statistics produced by national governments. The United States' portion of the cognitive interviewing, conducted by the National Center for Health Statistics, found that most respondents completely misunderstood the questions because they had no knowledge of the Federal Statistical System (Willson 2010). The respondents that did have some notion of the FSS were those who had made use of federal statistics (usually for work or educational purposes). Not surprisingly, respondents with first-hand experience with federal statistics had less difficulty making sense of the questions and had interpretations that were more consistent with question intent than those with no experience with government statistics were.

The current study produced results similar to the OECD project. We found that respondents have very little knowledge of federal statistics. As a result, the survey questions fall victim to a phenomenon common among many attitude questions; that is, there is no static underlying evaluation to capture. Instead, responses are created on the spot and are often inconsistent across questions aiming to measure similar concepts. In order for respondents to answer a survey question about a topic they have not previously considered, respondents take into account a variety of considerations about the issues and then make judgments about which considerations to use when answering the question. These considerations often consist of a random assortment of feelings, beliefs, impressions, and general values. The more heterogeneous these considerations, the more instability there is among responses. In other words, at any

given time, or among different questions designed to measure the same construct, respondents may sample and apply different considerations when formulating their answer. This process, referred to as the Belief Sampling Model (Tourangeau, Rips, Rasinski, 2000), explains response instability evident in many attitude questions. Furthermore, it is important to note that response instability is concerning to the extent that it indicates that a single concept is not being measured by a question. This diminishes our faith in the item's construct validity.

We found four forms of evidence supporting the notion that respondents do not have predefined opinions about federal statistics and that their answers to the survey questions lack stability. First, interpretations shifted among questions attempting to measure similar concepts. We also found that attitudes themselves shifted as respondents weighed different considerations on the topic. Second, in order to answer a question, respondents sometimes were not thinking about statistics at all, but rather drawing on other (often-irrelevant) considerations and topics. Third, sometimes there was general confusion over what a question was asking, to the point that respondents could not answer the question. Finally, respondents often limited their understanding of a question to the examples given, rather than thinking broadly about federal statistics. All of these factors suggest that respondents do not have specifically formulated underlying evaluations, which are knowable and measurable. The following discusses these four themes in more detail.

1. Shifting interpretations: One example of a lack of stability in the survey responses is the shifting interpretation apparent among respondents. There were many instances of respondents giving contradictory answers to similar survey questions or providing a narrative during probing that did not match the way they answered the survey question.

For example, one respondent disagreed with the statement “federal statistical agencies can get information collected by any one of them” because she believed the opposite to be true. She said, “I don’t think they share *enough* information. Unfortunately. You would think in this technology age they would, but they don’t.” However, in questions related to record linkage this same respondent said she was very much against her personal information being shared. This question is about sharing undefined, generic information. However, when she thinks a question is about sharing her personal information, her opinion is different. This lack of specificity in the question results in shifting interpretations as people sample different values and beliefs.

A similar phenomenon occurred for one respondent answering the question, “How do you feel about federal agencies collecting information directly?” At first, she said she supports “anything that protects my privacy.” However, in talking about accuracy, she says that if a record already exists, it would be more accurate than a survey. She discussed linkage to health records and income data, suggesting, “Their memory would be better than mine.” Again, this shifting attitude is clearly prompted by having the respondent think about different considerations surrounding a topic. Another respondent had the same issue in the question, “How do you feel about federal agencies collecting information directly?” He said, “I’d rather give my information directly – don’t go behind my back.” However, in a previous question he said he did not mind if the government gets information about him from various agencies. It seems clear that this question caused him to think about privacy differently from the way he thought about it in the previous question.

There were many examples of this pattern in the data. It is apparent that people’s answers easily fluctuate, depending on what they are prompted to think about and what considerations they sample before answering. This phenomenon is partly explained by the fact that broad topics are highly subject to context effects. This suggests that more stable estimates will be obtained by asking specific questions that cue all respondents to sample similar considerations. This idea will be discussed further in the section on proposed solutions.

2. *Not thinking about statistics:* Many people do not have predefined opinions about federal statistical data because they have little awareness of this type of information to begin with. Even those with some level of awareness do not possess sophisticated knowledge of statistical information and have not thought extensively about the topic, especially in relation to the Federal government. In order to answer the survey questions, respondents who fall into this category may sample from a set of considerations that are irrelevant to statistics. In essence, they are not understanding (or answering) the question as intended because they do not have the required knowledge base to do so. Additionally, this problem was exacerbated by some questions that were worse than others at conveying the idea that they were about statistics.

For example, one respondent agreed with the question “Statistics provided by federal agencies are often biased.” However, when asked why he agreed, he said that the lobby industry is influential. Lobbyists have the ability to persuade members of congress to vote certain ways and for certain policies. Similar misinterpretations were evident in the question “There is political interference in the work of federal statistical agencies.” To explain why she agreed with the statement, one respondent said, “I’ll just say one word: lobbyists!” She mentioned that politicians are influenced by lobbyists to vote in ways that sometimes contradict the platform on which they were voted in. These kinds of explanations illustrate how some respondents (those with no knowledge of federal statistics) do not interpret the questions as intended, but rather focus on what they are familiar with, which may be Congress instead of statistical agencies. As a result, the statistics produced by these items will not reflect the desired construct.

3. *Confusion over what the question was asking:* A related point is that respondents with very little knowledge of federal statistics sometimes had difficulty understanding what a question was asking altogether. If this confusion was great enough, they could not determine what beliefs to sample and, therefore, could not answer the question at all. For example, in the question “There is political interference in the work of federal statistical agencies,” one respondent could not answer because it did not make sense to him. He asked, “Who would be doing the interfering? Because I thought the government would be one big government, so who would be playing interference?”

Most times, however, confused respondents (i.e., those who were unaware of federal statistics and, therefore, could not understand the questions as intended) did give an answer. This created a variety of scenarios. Sometimes respondents could not stay on topic and had a difficult time understanding the nature of the questions. In these cases, survey administration was quite burdensome, for both the interviewers and the respondent. Other times respondents would draw upon their own experience in relation to the topic (not statistics on the topic). Several respondents answered the questions based on their own experiences with being laid off from work and trying to find a job. Finally, but less commonly, respondents would provide completely unrelated examples. For example, one respondent says that “If the Fed says they’re gonna do something, they’re gonna do it. If they say you’re going to be in jail for 6 years, you will be there for 6 years, not a day more or a day less.”

The question “Information collected to create federal statistics is sometimes used by the police and the FBI to keep track of people who break the law” is a good example of how respondents think about all kinds of information. Many respondents were not thinking of statistics as much as they were thinking about the government accessing peoples’ personal files or police records. They cited examples such as terrorists lists, sexual predator lists, lists of traffic ticket recipients, travel records (such as airline tickets), and personal files (one person talked about how they kept a large file on John Lennon).

4. *Interpretations limited to examples given:* A final indication that respondents cannot think generally about federal statistics is that many respondents limited their interpretation of the questions to the examples given in the first question, especially if they were already familiar with the agency. (The Census Bureau was, by far, the most recognized agency, but others were mentioned as well.) Even when

subsequent questions asked about federal statistics more generally, some respondents thought specifically of the agency they knew, such as Census.

This is not necessarily an unwelcomed pattern and, in fact, suggests a potential solution to the challenges of these questions. Respondents who lack a clear understanding of the FSS need something on which to base their answers. Rather than leave this up to the haphazard nature of heterogeneous belief sampling, it is preferable to have respondents at least think about a federal statistical agency when formulating their answers. The next section covers this and other possible improvements to the questions.

Proposed Solutions: Define the Context, Be Specific, Keep it Simple

Because most people in the general population have little or no knowledge of the FSS and have given little thought to statistical information in general, these attitude questions, like many attitude questions, are likely to produce unstable estimates. The challenge, then, for question design is to craft questions that elicit consistent interpretations among respondents. When piecing together an answer, it is important that respondents consider similar factors when arriving at an answer. For example, if when asked about their attitude regarding information sharing within the government, one group of respondents thinks about non-descript information being shared while another group thinks about personal identifiers, then the question is not measuring the same concept.

We suggest three question design strategies for improving construct validity. First, it is imperative to define concepts early and often. Respondents do not have shared understandings (or much understanding at all) of the concept of federal statistics; therefore, the questions must adequately define the context. Second, questions should be specific rather than general. Third, the question should be as simple as possible, in not only grammar and wording, but also conceptually. Multiple concepts should not exist in the same question. These strategies are discussed next.

1. *Define concepts up front:* Because many people do not have a good deal of knowledge about federal statistics, it is important that questions convey this topic right away and consistently. For example, before asking questions about people's understanding or use of "federal statistics," respondents were asked about specific statistics like the population count, unemployment rate, the poverty rate. This turned out to be the most important function of this series, and it could be strengthened in this regard by eliminating non-federal statistics. Rather than a test of knowledge (which we already know is limited), it should serve as a "primer" to define what it is we mean by federal statistics and federal statistical agencies.

2. *Be specific:* Many respondents in our sample were thinking of specific federal agencies when answering the questions. The most common agency cited was the Census Bureau, but others were mentioned too, such as the Bureau of Labor Statistics. Similarly, respondents tended to think specifically about unemployment statistics or the population count even when the question was asking them to think generally.

The consequence of honing in on specific examples is that the question is measuring opinions on those examples only. However, we contend that this is preferable to respondents not thinking about federal agencies or statistics at all – which tended to happen when they were presented with broad topics such as "federal statistics."

3. *Keep it simple:* Each question should be as straightforward as possible and avoid complex concepts that require higher-level thought and analysis on the part of respondents. Many respondents are not aware of nor have given much thought to federal statistics. Even those with higher awareness levels do not possess sophisticated thinking on the matter. Therefore, it is important to keep each question simple and to the point. Questions should not mix concepts or present complicated scenarios. Nor should they contain broad concepts that invite multiple understandings. If context needs to be provided, this can be accomplished by asking pointed and short questions such as "Have you ever heard about unemployment rate?" and "Do you happen to know who measures the unemployment rate?" By introducing several

examples of statistics in this way, early in the survey we were able to define federal statistics with a relatively short preamble: “Numbers like the unemployment rate, the population count and obesity statistics are statistics produced by agencies of the federal government. We call them ‘federal statistics.’”

Findings on Administrative Records Views

The questions on views towards administrative records had less definitive findings in this round of cognitive testing. Evidence from one question that had a rotating frame (some respondents saw one context first, while others saw another context first) suggests that questions on administrative records are vulnerable to context effects and shifting interpretations. Prior to taking this survey, respondents had given little thought to the topic. As a result, their opinions were swayed by the question wording and the topics presented therein. For example, we observed that respondents were more likely to agree than disagree with each version of the question shown below, even though they were intended to communicate two opposing viewpoints..

Sometimes federal statistical agencies need to get information such as employment history or retirement benefits. They can do it by getting the information from other government agencies or by asking people for it directly in a survey. Some people think people’s privacy would be better protected if each agency collected the information directly through surveys. How do you feel about federal agencies collecting information directly? *Are you strongly in favor, somewhat in favor, somewhat against, or strongly against?*

- a. How do you yourself feel about federal agencies trying to save government money and save people’s time by sharing information with each other? *Are you strongly in favor, somewhat in favor, somewhat against, strongly against?*

The first question asks if respondents support the government obtaining their information directly in a survey. A majority of respondents said they favor direct survey collection. Conversely, the second question (a) asks if they support the support record linkage over survey administration in order to save time and money, and a majority favored that as well. It is also worth mentioning that when respondents did NOT favor record linkage (asked a), most cited privacy concerns as the reason why. This is notable, as it was an issue arising from the respondents themselves and not suggested to them by the question. Privacy issues are related to the next point as well.

Questions on record linkage had some elements that were open to interpretation, which caused a degree of confusion. One example of wording that created confusion was the term “information.” Several respondents voiced this confusion when they discussed the idea of federal agencies obtaining information about them through record linkage versus through a survey. Respondents expressed concern over personal information being shared, which was predominantly defined as name, address, social security number, and the like. A couple of respondents mentioned information such as medical history or income as personal. The common theme here is privacy concerns. Vague terms like “collecting information directly” (what does directly mean?) also added to the confusion. For example, in the long question above, if respondents focused on the sentence, “How do you feel about federal agencies collecting information directly?,” they did not understand what “directly” meant. This sentence, when considered outside the context of the rest of the question, is not well defined. One person asked, “collected directly compared to what?” Another person said, “As opposed to indirectly?,” not understanding what it would mean to collect data indirectly.

We also have evidence to suggest that another source of confusion was long questions on administrative records. Some context is certainly needed. However, when there was too much information in the question, respondents lost track of what it was asking *or* were able to focus on only one part of the

question instead of the question in its entirety (thereby missing the intent and altering the desired construct).

Because administrative record linkage is not a topic to which most people have given much thought, questions should communicate intent as simply as possible and be specific in the information they have respondents consider. Without further testing at this point, we would suggest specifying the type of information being shared so that respondents do not have to speculate and specifically stating “ask people for it directly in a survey or...” in order to specify what “directly” actually means. Similarly, address privacy concerns while also being very specific about the type of information being collected.

These findings, as well as question-specific findings were incorporated into the questionnaire that was pretested for three weeks. Results from that pretest then were fielded for this survey. Appendix B shows the set of questions that went the field pretest described. The field test is described in detail in Earp, et al. (forthcoming).

5. Conclusions

The survey that was constructed based on the pretesting described here and in Earp, et al. (forthcoming), measures public knowledge, trust, credibility, confidence in institutions and views towards administrative records. It is the baseline survey that began data collection on February 15, 2012. From this date, until September 30, 2013, we will gather trend data on how the American people view the FSS. We will be able to see the impact of current events or public relations campaigns on knowledge, trust, credibility and attitudes towards administrative records. The cognitive testing described here and in Willson (2011) allowed us to craft a questionnaire that helped to guide respondents into a mindset in which they could answer questions about an otherwise vague topic. By grounding them with statistics that they were more likely to have heard of and by describing what was meant by “federal statistics” we felt confident that the set of questions that followed were measuring what was intended to be measured.

After this round of pretesting, the team felt much more confident in the questions on knowledge and trust and credibility than those on administrative records views. As such, we launched another research agenda to further develop questions on administrative records views that will be incorporated into the Gallup survey using the rotating questions feature. To date, we have conducted focus groups and additional cognitive interviews in support of this objective.

Future papers and presentations will detail findings from this survey. This paper serves to set the backdrop for the survey and demonstrate its conception, how it was developed and qualitatively pretested. Earp et al (forthcoming) will demonstrate the quantitative pretest and preliminary findings.

Acknowledgements

We wish to acknowledge the members of the ICSP who commissioned this survey, the Gallup organization who is fielding this survey, as well as Eleanor Singer, Peter Miller, Michelle Smirnova, Paul Scanlon, and Brian Harris-Kojetin, who have been critical in executing this work. We also thank Terry DeMaio and Nancy Bates for reviewing this paper and providing helpful feedback.

References

- Brackfield, D. (2011.) “OECD Work on Measuring Trust in Official Statistics”
- Conrey, F.R., ZuWallack, R., and Locke, R. (2011). Census Barriers, Attitudes and Motivators Survey II Final Report, 2010 Census Program for Evaluations and Experiments, October 31, 2011.
- Curtin, R. (2007), “What US Consumers Know About Economic Conditions”, Paper presented at the second OECD Workshop on “Measuring and Fostering the Progress of Societies, Istanbul, June 27.

Fellegi, I. (1996). Characteristics of an effective statistical system. *Canadian Public Administration*, Vol 39, Issue 1, pg. 5-34.

Fellegi, I. (2004). “Maintaining the Credibility of Official Statistics” in *Statistical Journal of the United Nations ECE* (21) 191–198.

Fellegi, I. (2010.) “Report of the electronic working group on measuring trust in official statistics”, *OECD Meeting of the Committee on Statistics*, June 2010, Paris.

Miller, P.V., Walejko, G. K. (2010). Tracking Study of Attitudes and Intentions to Participate in the 2010 Census. Technical Report submitted to the U.S. Census Bureau for Contract YA1323-09-CQ-0032, Task Order 001, December 10, 2010.

OECD Working Group. (2011.) “Measuring Trust in Official Statistics—Cognitive Testing”

ONS (2009) Data-sharing Between Government Departments. A Report on Public Acceptability. Newport: Office of National Statistics.

Pascale, J. (2011). “Requesting Consent to Link Survey Data to Administrative Records”. A paper presented at the Fourth Conference of the European Survey Research Association (ESRA), Lausanne, Switzerland, July 18-22, 2011.

Singer, E., Bates, N. & Van Hoewyk, J. (2011). Concerns about privacy, trust in government, and willingness to use administrative records to improve the decennial census. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Phoenix Arizona.

Willson, S. (2010). Cognitive Interview Evaluation of Survey Questions Measuring Trust in Official Statistics. Questionnaire Design Research Laboratory, National Center for Health Statistics, Centers for Disease Control and Prevention. Internal Report.

Willson, S. (2012). Cognitive Interview Evaluation of the Federal Statistical System Trust Monitoring Survey: Results of interviews conducted in October, 2011. Questionnaire Design Research Laboratory, National Center for Health Statistics, Centers for Disease Control and Prevention. Internal Report.

Wroblewski, M., Bates, N., and Pascale, J. (2012). Public Attitudes toward the Use of Administrative Records in the U.S. Census: Does Question Frame Matter? Federal Committee on Statistical Methodology Research Conference. Washington, DC.

Appendix A: Cognitive Testing Questions Mapped to Fellegi Constructs

1. Please tell me if you have ever heard or read about the following numbers on the radio, TV, newspapers, the internet, or anywhere else. (*Adapted from Fellegi, 2010; awareness of statistics*)
 - a. Have you ever heard or read about the unemployment rate?
 - i. *If yes*, Do you happen to know what organization measures the unemployment rate?
 1. Dept. of Labor
 2. Bureau of Labor Statistics
 3. Census Bureau
 4. Federal Government
 5. Media
 6. Other, specify _____
 - b. Have you ever heard or read about the total number of people in the United States or the population count?
 - i. *If yes*, Do you happen to know what organization conducts the population count?
 1. Dept. of Commerce
 2. Census Bureau
 3. Federal Government

4. Media
 5. Other, specify _____
 - c. Have you ever heard or read about obesity statistics?
 - i. *If yes*, Do you happen to know what organization measures obesity?
 1. Dept. of Health and Human Services
 2. Centers for Disease Control
 3. National Institute of Health
 4. National Center for Health Statistics
 5. Hospitals
 6. Federal Government
 7. Media
 8. Other, specify _____
 - d. Have you ever heard or read about the Neilson TV ratings?
 - i. *If yes*, Do you happen to know who calculates the Neilson TV Ratings?
 1. Neilson
 2. Federal Government
 3. Media
 4. Other, specify _____
2. When important decisions need to be made based on statistics, which of the following sources is more believable to you: (*Integrity*)
- a. A University
 - b. An agency of the Federal government
 - c. A private company
 - d. A political party
 - e. The Media
3. Numbers like the unemployment rate, the population count and obesity statistics are statistics produced by agencies of the federal government. We call them “federal statistics.” Have you ever *used* or talked about federal statistics like the unemployment rate, the population count or obesity statistics for study, work, or any other purpose? (*Yes, No, DK*)
- a. (*If yes*): Have you used federal statistics frequently, occasionally, or only once or twice? (*Adapted from Fellegi, ibid.; measures personal relevance*)
 - b. No
 - c. Don’t Know
4. To what extent do you agree or disagree with the following statements about federal statistics?
1. Federal statistics on unemployment, population, and health are important for understanding our society. *Do you strongly agree, somewhat agree, somewhat disagree, or strongly disagree?* (*adapted from OECD, relevance*)
 2. Policy makers need federal statistics to make good decisions about things like federal funding. (*relevance*)
 3. State and local government officials need federal statistics to make good decisions about things like where to locate hospitals and schools. (*Relevance*)
 4. Statistics provided by the federal agencies are generally accurate. *Do you agree strongly, etc.* (*adapted from ONS, accuracy*)
 5. Federal statistics give a good picture of life in the United States. (*accuracy*)
 6. Statistics provided by federal agencies are often biased. (*objectivity*)
 7. Statistics produced by federal agencies, like the Census Bureau and the Bureau of Labor Statistics, do not favor one political party or another. (*impartiality*)
 8. There is political interference in the work of federal statistical agencies. (*impartiality*)
 9. People can trust federal statistical agencies to keep information about them confidential. (*OECD, confidentiality protected*)
 10. Federal statistical agencies share too much information with each other. (*confidentiality protection*)

11. All federal statistical agencies can get information collected by any one of them. (*confidentiality protection*)
 12. Information collected to create federal statistics is sometimes used by the police and the FBI to keep track of people who break the law. (*Prior Census Bureau study, confidentiality protected*)
 13. Federal statistical agencies give personal information about people to marketing firms. (*confidentiality protection*)
 14. If I needed to, I could easily find out exactly how federal statistics are produced. (*transparency*)
 15. Federal statistical agencies are honest and professional. (*integrity*)
 16. Private companies could produce more accurate statistics than Federal statistical agencies. (*integrity*)
5. Now I'm going to read you a list of organizations in American society. Please tell me how much confidence you, yourself, have in each one – a great deal, quite a lot, some or very little? (*adapted Gallup, confidence in institutions*)
- a. The mass media, such as newspapers, radio, and television.
 - b. Bloggers on the Internet.
 - c. The Federal government.
 - d. Federal statistics, such as the unemployment rate, the population count, or obesity statistics.
 - e. Political polls.
 - f. Your state government.
 - g. Banks.
 - h. Large corporations
 - i. The U.S. Supreme Court.

Appendix B: Field Tested Questions Mapped to Fellegi Constructs

Fellegi Construct	Wording	Week 1	Week 2	Week 3
	Now I have some questions on a different topic. At a few points throughout this next series, I will ask you to tell me why you chose your answer to that question. This will be helpful information as we try to understand how different people interpret our questions.	x		
	I will read you some numbers that you may have heard of or read about on the radio, TV, newspapers, the Internet or somewhere else. Please tell me if you have heard of them:	x		
Awareness	The Unemployment rate	x		
Awareness	The total number of people in the United States, or the population count	x		
Awareness	The poverty rate	x		
Awareness	(You mentioned that you have heard of the unemployment rate.) Do you happen to know who measures the unemployment rate (in the U.S.)?	x	x	
Awareness	(You mentioned that you have heard of the population count of the total number of people in the United States.) Do you happen to know who conducts the population count (who measures the total number of people in the U.S., or the population count)?	x	x	x
Awareness	(You mentioned that you had heard of the poverty rate.) Do you happen to know who measures the poverty rate (in the U.S.)?	x	x	
Awareness	Do you happen to know who measures the number of deaths in the U.S. caused by different diseases?		x	x
Awareness	Do you happen to know who measures the crime rate in the U.S.?		x	x

Awareness	Do you happen to know who measures the U.S. Consumer Price Index or CPI ?		x	x
Relevance	Numbers like the unemployment rate, the population count and the poverty rate are federal statistics produced by federal statistical agencies that are part of the federal government. Have you ever used federal statistics for study or work?	x	x	x
Integrity	When important decisions need to be made based on statistics, which of the following sources is the most believable to you: A University, An agency of the Federal Government, A Private Company, A Political Party, The Media?	x	x	x
Credibility	Personally, how much trust do you have in the federal statistics in the United States? Would you say that you tend to trust federal statistics or tend not to trust them?	x	x	x
Relevance	Policy makers need federal statistics to make good decisions about things like federal funding.	x	x	x
Relevance	State and local government officials need federal statistics to make good decisions.	x	x	
Accuracy	Statistics provided by the federal agencies are generally accurate.	x	x	x
Accuracy	The unemployment rate gives a true picture of what is happening to unemployment.	x	x	
Objectivity	Statistics provided by federal agencies are often biased.	x	x	x
Impartiality	Elected officials interfere with the production of statistics by federal agencies.	x	x	x
Impartiality	Federal statistics are made public only if approved by the president or congress.	x	x	
Confidentiality protected	People can trust federal statistical agencies to keep information about them confidential.	x	x	x
Confidentiality protected	Federal statistical agencies give personal information about people to the IRS.	x	x	
Confidentiality protected	Federal statistical agencies give personal information about people to marketing firms.	x	x	
Transparency	People can easily find out exactly how federal statistics are produced.	x	x	x
Integrity	Federal statistical agencies are honest.	x	x	x
Integrity	Federal statistical agencies have the experts they need to produce high quality statistics. (Federal statistical agencies are a reliable source for high quality statistics.)	x	x	x
Integrity	Private companies could produce more accurate statistics than Federal statistical agencies.	x	x	
	Now I'm going to read you a list of organizations in American society. Please tell me how much confidence you, yourself, have in each one – a great deal, quite a lot, some or very little?			
	Newspapers	x	x	x
	The military	x	x	x
Integrity	Federal statistical agencies	x	x	x
	Congress	x	x	x